Recent Developments in Statistics and Actuarial Science

A Festschrift in Celebration of Professor M.M. Ali's Twenty-five Years at Western

Editors:
M. Safiul Haq
Serge B. Provost

Managing Editor:
Maria I. Hlawka-Lavdas

25

SciTEX Publications
The University of Western Ontario

Ram C. Tiwari,[1] Siddhartha Chib[2] and S. Rao Jammalamadaka[3]

NONPARAMETRIC BAYES ESTIMATION OF MULTIVARIATE REGRESSION WITH A DIRICHLET INVARIANT PRIOR

ABSTRACT

The paper considers nonparametric Bayes estimation of the regression coefficient matrix in a multivariate regression problem under the squared error loss function using a Dirichlet invariant process prior, and a mixture of Dirichlet invariant process priors. This generalizes the work of Poli (1985).

1. INTRODUCTION

Consider a multivariate regression problem where we have a sample of n independent observations $\{y_{1i}, \ldots, y_{pi}\,;\ x_{1i}, \ldots, x_{qi}\}$, $i = 1, \ldots, n$, on a set of $d\,(= p+q)$ random variables $\boldsymbol{z} = (\boldsymbol{y}', \boldsymbol{x}')' \in R^d$, with $\boldsymbol{y} \in R^p$ and $\boldsymbol{x} \in R^q$ having the joint distribution F which is assumed to be unknown. We are primarily interested in capturing the relationship between the variables $\boldsymbol{y}$ and $\boldsymbol{x}$.

Data of this type is often analyzed using the multivariate regression model in which we assume the linear relation

$$Y = XB + E \tag{1.1}$$

[1] Department of Mathematics, University of North Carolina, Charlotte, North Carolina 28223

[2] Department of Economics, University of Missouri-Columbia, Columbia, Missouri 65211

[3] Statistics and Applied Probability Program, University of California, Santa Barbara, California 93106

where Y is an $n \times p$ matrix whose ith row is $\boldsymbol{y}_i' = (y_{1i}, \ldots, y_{pi})$, X is an $n \times q$ matrix with ith row $\boldsymbol{x}_i' = (x_{1i}, \ldots, x_{qi})$ and E is a random $n \times p$ matrix with mean 0 and covariance matrix $\Sigma \otimes I$ for some $p \times p$ positive-definite matrix Σ. Focus is on estimating the $q \times p$ parameter matrix B. Under these assumptions the ordinary least squares estimator of B is given by

$$\hat{B} = (X'X)^{-1}X'Y \tag{1.2}$$

which does not depend on Σ (see, e.g., Press (1982), p. 231–232).

The corresponding parametric Bayes analysis of the model (1.1), usually under the assumption that E is matrix variate normal, proceeds by assuming a prior distribution for B and Σ and finding the posterior of B through an application of the Bayes rule. The details are available in Zellner (1971) and Press (1982).

In the nonparametric Bayes framework, as exposited by Goldstein (1976) (also, see Prakasa Rao (1983), Section 11.4) and employed in Poli (1985), model (1.1) is not assumed to hold. Instead, a predictive approach is adopted in which the matrix B is chosen to minimize the mean-squared prediction error

$$E^{F|Z}\left[\left(\boldsymbol{y}_{n+1}' - \boldsymbol{x}_{n+1}'B\right) W \left(\boldsymbol{y}_{n+1}' - \boldsymbol{x}_{n+1}'B\right)'\right] \tag{1.3}$$

where $\boldsymbol{y}_{n+1}' = (y_{1,n+1}, \ldots, y_{p,n+1})$ and $\boldsymbol{x}_{n+1}' = (x_{1,n+1}, \ldots, x_{q,n+1})$ are the $(n+1)st$ future observations from F, W is a weight matrix, and $E^{F|Z}$ denotes the expectation with respect to the distribution of $\boldsymbol{z}_{n+1} = (\boldsymbol{y}_{n+1}', \boldsymbol{x}_{n+1}')'$ given the data $Z = [Y : X]$; that is, the predictive distribution. The minimizing solution to (1.3) is given by

$$B^* = D^{-1}A \tag{1.4}$$

where $D = E^{F|Z}(\boldsymbol{x}_{n+1}\boldsymbol{x}_{n+1}')$ is of order $q \times q$ and is assumed nonsingular, and $A = E^{F|Z}(\boldsymbol{x}_{n+1}y_{n+1}')$ is of order $q \times p$.

As stated in Poli (1985), "the achieved estimate B^* provides the best linear prediction of $\boldsymbol{y}_{n+1}$ in terms of $\boldsymbol{x}_{n+1}$, without assuming however that the conditional mean of $y_{i,n+1} (i = 1, \ldots, p)$ is a linear function of x_{n+1}."

In order to compute (1.4), we need to find the predictive distribution function of $\boldsymbol{z}_{n+1}$ given the data. If we let

$$F(\boldsymbol{z}) = P(-\infty, \boldsymbol{z}] \tag{1.5}$$

where P is a probability measure on $(R^d, \boldsymbol{R}^d)$ with the distribution F, then the predictive distribution of $\boldsymbol{z}_{n+1}$ given the data is obtained in the Bayes setup by assigning a prior distribution on the space P of all probability measures, then deriving the posterior distribution of F given the data Z and computing its mean.

An analysis along these lines is conducted in Poli (1985) under the assumption that the prior distribution of F is (i) Ferguson's (1973) Dirichlet process prior, and (ii) Antoniak's (1974) mixture of Dirichlet process priors. Both these priors are defined on $(P, \sigma(P))$, where $\sigma(P)$ is the smallest σ-field of subsets of P such that the map $P \to P(A)$ is $\sigma(P)$-measurable for each A in $\boldsymbol{R}^d$.

In this note, we generalize Poli's (1985) work by assuming that prior beliefs are represented by Dalal's (1979) Dirichlet invariant process prior which yields posterior distributions that are invariant under a given group of transformations. It is noteworthy that with this formulation we can, for example, assign priors on the class of distributions that are symmetric around arbitrary points, or on the class of distributions that are exchangeable in coordinates.

The paper is organized as follows. In Section 2, some definitions and results on the Dirichlet σ invariant process priors are presented. They are used in the sequel. In Section 3, the Bayes estimator B^* of the regression coefficient matrix B is derived under a Dirichlet invariant process prior, and then under a mixture of Dirichlet invariant process priors.

2. PRELIMINARIES ON A DIRICHLET INVARIANT PROCESS PRIOR

In this section, we introduce some definitions and results that are used in the sequel. Let (Ω, S, Q) be a probability space.

Definition 2.1. A random probability measure on P on $(R^d, \boldsymbol{R}^d)$ is a measurable map from (Ω, S, Q) into $(P, \sigma(P))$. The induced measure QP^{-1} on $(P, \sigma(P))$ is called the prior distribution of P.

A random probability measure P can be regarded as a transition function $P(\cdot,\cdot)$ from $\Omega \times \boldsymbol{R}^d \to [0,1]$ such that, for each ω in Ω, $P(\omega,\cdot)$ is a probability measure on $(R^d, \boldsymbol{R}^d)$ and, for each A in R^d, $P(\cdot, A)$ is an S-measurable function. If P is a random probability measure and F is given by (1.5), then F is called a random distribution function.

Let $G = \{g_1, \ldots g_k\}$ be a finite group of measurable transformations from R^d into itself. A measure ν is said to be G-invariant if $\nu(gA) = \nu(A)$ for all $g \in G$ and all $A \in \boldsymbol{R}^d$. A set $A \in \boldsymbol{R}^d$ is called G-invariant if $gA = A$ for all $g \in G$. By a G-invariant measurable partition $(A_1, \ldots, A_r)$ of R^d, we mean that the sets of the partition are G-invariant and $\boldsymbol{R}^d$-measurable.

Definition 2.2. (Dalal (1979)). Let α be a finite, nonnull G-invariant measure on $(R^d, \boldsymbol{R}^d)$ and let P be a G-invariant random probability measure on $(R^d, \boldsymbol{R}^d)$. We say P has the Dirichlet invariant process prior $DG(\alpha)$ with parameter α, if, for every finite G-invariant measurable partition $(A_1, \ldots A_r)$ of R^d, the joint distribution of $(P(A_1), \ldots, P(A_r))$ under $DG(\alpha)$ is the singular Dirichlet distribution $D(\alpha(A_1), \ldots, \alpha(A_r))$ as defined in Wilks (1962).

If F is the random distribution function associated with a random probability measure P on $(R^d, \boldsymbol{R}^d)$, we use the notation $F \in DG(\alpha)$ to denote that the prior distribution of P is $DG(\alpha)$. When G consists of a single element, namely the identity transformation, the Dirichlet-invariant process $DG(\alpha)$ is indeed Ferguson's (1973) Dirichlet process prior $D(\alpha)$ with parameter α. Alternative definitions of the Dirichlet invariant process prior are given by Dalal (1975) and Tiwari (1981).

Let $F \in DG(\alpha)$. By a random sample $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ of size n from F we mean that, given F, the random vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ are independent and identically distributed with common distribution F. For any $\boldsymbol{z} \in R^d$, let $\delta_{\boldsymbol{z}}$ denote the degenerate probability measure at $\boldsymbol{z}$.

Theorem 2.3. (Dalal (1975)). Let $F \in DG(\alpha)$ and let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ be a random sample of size n from F, then the conditional distribution of the random probability measure P (associated with F) is the Dirichlet invariant process prior with the (updated) parameter

$$\alpha + \sum_{j=1}^{n} k^{-1} \sum_{i=j}^{k} \delta_{g_i(\boldsymbol{z}_j)}; \quad \text{that is,} \quad F \mid Z \in DG\left(\alpha + \sum_{j=1}^{n} k^{-1} \sum_{i=1}^{k} \delta_{g_i(\boldsymbol{z}_j)}\right).$$

Let $G_{\boldsymbol{\mu}} = \{e, g_{\boldsymbol{\mu}}\}$ where $e(\boldsymbol{z}) = \boldsymbol{z}$ and $g_{\boldsymbol{\mu}}(\boldsymbol{z}) = 2\boldsymbol{\mu} - \boldsymbol{z}$ for $\boldsymbol{z} \in R^d$ and $\boldsymbol{\mu} \in R^d$. Let $\alpha_{\boldsymbol{\mu}}$ be symmetric about $\boldsymbol{\mu}$, i.e. $g_{\boldsymbol{\mu}}$-invariant. Then we have the following:

Theorem 2.4. (Dalal (1975)). Let $F_{\mu} \in DG(\alpha_{\mu})$ and, given μ and F_{μ}, let $Z = (z_1, \ldots, z_n)$ be a random sample from F_{μ}. Let μ have prior density ξ. Then

$$F \mid Z, \mu \ \in DG\Big(\alpha_{\mu} + \frac{1}{2}\sum_{j=1}^{n}(\delta_{z_j} + \delta_{2\mu - z_j})\Big)\ ,$$

and

$$F \mid Z \in \int_{R^d} DG\Big(\alpha_{\mu} + \frac{1}{2}\sum_{j=1}^{n}(\delta_{z_j} + \delta_{2\mu - z_j})\Big) d\xi(\mu \mid Z)\ .$$

when $\xi(\cdot \mid Z)$ is the density of μ given the data.

From Theorem 2.3, the predictive distribution of z_{n+1} given the data Z is

$$\begin{aligned}\hat{F}(t) &= E^{F|Z}[F(t)] \\ &= p_n\bar{\alpha}(t) + (1 - p_n)\frac{1}{nk}\sum_{j=1}^{n}\sum_{i=1}^{k}\delta_{g_i(z_j)}(-\infty, t]\ ,\end{aligned} \qquad (2.1)$$

where

$$p_n = M/(M + n),\quad \bar{\alpha}(t) = \alpha(-\infty, t]/M,\quad M = \alpha(R^d)\ . \qquad (2.2)$$

In the case when $G_{\mu} = \{e, g_{\mu}\}$ with $e(z) = z$, $g_{\mu}(z) = 2\mu - z$, $z \in R^d$, from Theorem 2.4 the predictive distribution of z_{n+1} is given by

$$\begin{aligned}\hat{F}(t) &= E^{\mu|Z}E^{F|Z,\mu}\{F(t)\} \\ &= E^{\mu|Z}\Big\{p_{n,\mu}\bar{\alpha}_{\mu}(t) + (1 - p_{n,\mu}) \\ &\qquad \times\frac{1}{2n}\sum_{j=1}^{n}[\delta_{z_j}(-\infty, t] + \delta_{2\mu - z_j}(-\infty, t]]\Big\}\end{aligned} \qquad (2.3)$$

where

$$p_{n,\mu} = \frac{M_{\mu}}{(M_{\mu} + n)}\ ,\quad \bar{\alpha}_{\mu}(t) = \frac{\alpha_{\mu}(-\infty, t]}{M_{\mu}}\ ,\quad M_{\mu} = \alpha_{\mu}(R^d)\ . \qquad (2.4)$$

The definition of a mixture of Dirichlet invariant process priors is given next. Let $(\Theta, \boldsymbol{A})$ and $(U, \boldsymbol{B})$ be two measurable spaces. A transition measure β is a mapping from $U \times \boldsymbol{A}$ into $[0, \infty)$ such that for every $u \in U$, $\beta(u, \cdot)$ is a finite, nonnegative, nonnull measure on $(\Theta, \boldsymbol{A})$, and for every $A \in \boldsymbol{A}$, $\beta(\cdot, A)$ is $\boldsymbol{B}$-measurable function. Note that this differs from the definition of a transition probability in that $\beta(u, \Theta)$ need not identically be one. We say a transition measure β is a G-invariant if for each $u \in U$, $g \in G$ and $A \in \boldsymbol{R}^d$, $\beta(u, gA) = \beta(u, A)$.

Definition 2.5. Let $(U, \boldsymbol{B}, H)$ be a probability space, called the index space, and let α be a G-invariant transition measure function on $U \times \boldsymbol{R}^d$. We say a G-invariant random probability measure P on $(R^d, \boldsymbol{R}^d)$ has the mixture of Dirichlet invariant process priors with mixing distribution H on $(U, \boldsymbol{B})$, and G-invariant transition measure α, if, given $u \in U$, P has prior $DG(\alpha(u, \cdot))$, where u has distribution H.

In concise symbols we use the notation:

$$F \in \int_U DG(\alpha(u, \cdot)) dH(u) ,$$

where F is the random distribution associated with P. When G has only a single element, a mixture of Dirichlet invariant processes reduces to Antoniak's (1974) mixture of Dirichlet processes. As an example, let P have prior $DG(\alpha)$. Define

$$\alpha(\boldsymbol{z}, A) = \alpha(A) + \frac{1}{k} \sum_{i=1}^{k} \delta_{g_i(\boldsymbol{z})}(A) .$$

Note that $\alpha(\boldsymbol{z}, gA) = \alpha(\boldsymbol{z}, A)$ for all $\boldsymbol{z} \in R^d$, $g \in G$ and $A \in \boldsymbol{R}^d$. Let H be a fixed probability measure on $(R^d, \boldsymbol{R}^d)$. Then the process P^* which chooses $\boldsymbol{z}$ according to H, and P from $DG(\alpha, \boldsymbol{z}, \cdot))$ is a mixture of Dirichlet invariant process priors as defined above. Moreover, if $(B_1, \ldots, B_r)$ is any G-invariant measurable partition of R^d, then

$$(P(B_1), \ldots, P(B_r)) \in \sum_{i=1}^{r} H(B_i) DG(\alpha(B_1), \ldots, \alpha(B_i) + 1, \ldots, \alpha(B_r)).$$

The properties of a mixture of Dirichlet invariant processes can be derived similarly to those of a mixture of Dirichlet processes (cf. Antoniak, 1974). For example, if

$$F \in \int_U DG(\alpha(u,\cdot))dH(u), \quad \text{and} \quad \boldsymbol{z} \mid F \sim F\ ,$$

then the marginal distribution of $\boldsymbol{z}$ is given by

$$P(\boldsymbol{z} \in A) = \int_U \bar{\alpha}(u,A)dH(u), \quad A \in \boldsymbol{R}^d\ , \tag{2.5}$$

where

$$\bar{\alpha}(u,A) = \alpha(u,A)/\alpha(u,R^d)\ , \tag{2.6}$$

since

$$\begin{aligned} P(\boldsymbol{z} \in A \mid u) =& E^H[P(\boldsymbol{z} \in A \mid u, P) \mid u] \\ =& E[P(A) \mid u] \\ =& \bar{\alpha}(u,A) \ \ \text{a.s.}[H] \end{aligned}$$

and hence,

$$P(\boldsymbol{z} \in A) = E^H \bar{\alpha}(u,A) = \int_U \bar{\alpha}(u,A)dH(u)\ .$$

Also, if $F \in DG(\alpha)$ and $\boldsymbol{z}$ is a random sample of size one from F, and if $A \in \boldsymbol{R}^d$, then

$$F \mid_{\boldsymbol{z} \in A} \in \int_A DG(\alpha(\boldsymbol{z},.))d\bar{\alpha}_A(\boldsymbol{z})\ ,$$

where

$$\alpha(\boldsymbol{z},\cdot) = \alpha + \frac{1}{k}\sum_{i=1}^{k} \delta_{g_i(\boldsymbol{z})} \text{ for } \boldsymbol{z} \in A\ , \text{ and } \bar{\alpha}_A(B) = \bar{\alpha}(B \cap A)/\bar{\alpha}(A)$$

for $B \in \boldsymbol{R}^d$. That is, the posterior distribution of F, given that $\boldsymbol{z}$ has fallen in a set A, is a mixture of Dirichlet invariant processes with an index space $(A, \boldsymbol{R}^d \cap A)$ and transition measure α on $A \times (\boldsymbol{R}^d \cap A)$, with mixing distribution $H_A = \bar{\alpha}_A$. Note that if $\boldsymbol{z}$ itself is observed, the conditional distribution of F given $\boldsymbol{z}$ is not a mixture but simply

$$DG\left(\alpha + \frac{1}{k}\sum_{i=1}^{k}\delta_{g_i(\boldsymbol{z})}\right),$$

and by induction on the sample size,

$$F \mid Z \in DG\left(\alpha + \sum_{j=1}^{n} k^{-1}\sum_{i=1}^{k}\delta_{g_i(\boldsymbol{z}_j)}\right).$$

Also, note that with $A = R^d$ we have $\int_{R^d} DG(\alpha + \frac{1}{k}\Sigma\delta_{g_i(\boldsymbol{z})})d\bar{\alpha}(\boldsymbol{z}) = DG(\alpha)$. Thus the conditional distribution of a Dirichlet invariant process given only the information that a sample was observed, and not its value, is the same as the original distribution. Alternatively, if we treated mixing as an operator, this would be the identity operator. Let (U, B) be a standard Borel space (see, e.g., Parthasarathy (1967)). Then from the above discussions, we have the following result.

Theorem 2.5. Let $F \in \int_U DG(\alpha(u,\cdot))dH(u)$. Given F, let $Z = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$ be a sample of size n from F. Then,

$$F \mid Z \in \int_U DG(\alpha(u,\cdot) + \sum_{j=1}^{n} k^{-1}\sum_{i=1}^{k}\delta_{g_i(\boldsymbol{z}_j)})dH(u \mid Z).$$

Now, from (2.5) and Theorem 2.5 the predictive distribution of $\mathbf{z}_{n+1}$ given the data Z under the assumption that $F \in \int_U DG(\alpha(u,\cdot))dH(u)$ is given by

$$\begin{aligned}\hat{F}(\boldsymbol{t}) &= E^{u|Z}E^{F|Z,u}\{F(\mathbf{t})\}\\ &= E^{u|Z}\left[p_n\bar{\alpha}(u,\boldsymbol{t}] + \frac{(1-p_n)}{nk}\sum_{j=1}^{n}\sum_{i=1}^{k}\delta_{g_i(\boldsymbol{z}_j)}(-\infty,\boldsymbol{t}]\right]\\ &= \int_U\left\{p_n\bar{\alpha}(u,\boldsymbol{t}] + \frac{(1-p_n)}{nk}\sum_{j=1}^{n}\sum_{i=1}^{k}\delta_{g_i(\boldsymbol{z}_j)}(-\infty,\boldsymbol{t}]\right\}dH(u \mid Z)\end{aligned} \tag{2.7}$$

where $\bar{\alpha}(u,\boldsymbol{t}] = \bar{\alpha}(u,(-\infty,\boldsymbol{t}])$ is given by (2.6), and $H(u \mid Z)$ is the posterior distribution of u given the data Z.

3. BAYES ESTIMATION

3.1 Bayes estimation with a Dirichlet invariant prior

Assume $G = \{g_1, \ldots, g_k\}$ is a finite group of transformations defined from R into itself. For any $z \in R^d$, define $gz = (gz_1, \ldots, gz_d)\ \forall g \in G$. Let (i, i')th element of matrix D in (1.4) be denoted $D_{ii'}$ and let the (i, j)th element of matrix A in (1.4) be denoted A_{ij}. Then under prior $DG(\alpha)$, and the assumption that

$$E^{\bar{\alpha}}(x_1^2) = \int_{R^d} x_i^2 d\bar{\alpha}(y, x) < \infty,$$
$$E^{\bar{\alpha}}(y_j^2) = \int_{R^d} y_j^2 d\bar{\alpha}(y, x) < \infty, \quad 1 \le i \le q, \quad 1 \le j \le p, \quad (3.1)$$

we have

$$\begin{aligned} D_{ii'} &= E^{F|Z}(x_{i,n+1} \cdot x_{i',n+1}) \\ &= p_n E^{\bar{\alpha}}(x_{i,n+1} \cdot x_{i',n+1}) + (1 - p_n) \frac{1}{nk} \sum_{m=1}^{n} \sum_{\ell=1}^{k} g_\ell(x_{im} \cdot x_{i'm}), \\ &\qquad i, i' = 1, \ldots, q, \end{aligned} \quad (3.2)$$

and

$$\begin{aligned} A_{ij} &= E^{F|Z}(x_{i,n+1} \cdot y_{j,n+1}) \\ &= p_n E^{\bar{\alpha}}(x_{i,n+1} \cdot y_{j,n+1}) + (1 - p_n) \frac{1}{nk} \sum_{m=1}^{n} \sum_{\ell=1}^{k} g_\ell(x_{im} \cdot y_{jm}) \\ &\qquad i = 1, \ldots, q, \quad j = 1, \ldots, p. \end{aligned} \quad (3.3)$$

From (3.2), the matrix D is

$$D = p_n E^{\bar{\alpha}}(\boldsymbol{x}'_{n+1} \boldsymbol{x}_{n+1}) + \frac{(1 - p_n)}{nk} \sum_{\ell=1}^{k} g_\ell(X'X), \quad (3.4)$$

and from (3.3) the matrix A is

$$A = p_n E^{\bar{\alpha}}(\boldsymbol{x}'_{n+1} \boldsymbol{x}_{n+1}) + \frac{(1 - p_n)}{nk} \sum_{\ell=1}^{k} g_\ell(X'Y). \quad (3.5)$$

Thus from (3.4) and (3.5) the Bayes estimator of B under prior $DG(\alpha)$ is given by

$$\begin{aligned} B^{*\alpha} &= D^{-1}A \\ &= \left\{ p_n E^{\bar{\alpha}} \left(\boldsymbol{x}_{n+1}\boldsymbol{x}'_{n+1} \right) + (1-p_n)\frac{1}{nk}\sum_{\ell=1}^{k} g_\ell(X'X) \right\}^{-1} \\ &\quad \times \left(p_n E^{\bar{\alpha}}(\boldsymbol{x}_{n+1}\boldsymbol{y}'_{n+1}) + (1-p_n)\frac{1}{nk}\sum_{\ell=1}^{k} g_\ell(X'Y) \right) . \end{aligned} \quad (3.6)$$

Note that (3.6) can be rewritten as

$$B^{*\alpha} = D^{-1}p_n \ E^{\bar{\alpha}} \left(\boldsymbol{x}_{n+1}\boldsymbol{x}'_{n+1} \right) B^0 + D^{-1}(1-p_n)\frac{1}{nk}\sum_{\ell=1}^{n} g_\ell(X'Y) \ , \quad (3.7)$$

where

$$B^0 = \left(E^{\bar{\alpha}} \left(\boldsymbol{x}_{n+1}\boldsymbol{x}'_{n+1} \right) \right)^{-1} E^{\bar{\alpha}} \left(\boldsymbol{x}_{n+1}\boldsymbol{y}'_{n+1} \right) \quad (3.8)$$

is the Bayes estimator of B for zero sample size. If we consider a sequence of Dirichlet invariant process priors $\{DG(\alpha_t)\}_{t=0}^{\infty}$ such that

$$\alpha_t(R^d) \to 0 \quad \text{and} \quad \sup_{A \in R^d} | \bar{\alpha}_t(A) - \bar{\alpha}_0(A) | \to 0 \ \text{ as } \ t \to \infty \ ,$$

then (see, e.g., Sethuraman and Tiwari (1982)) under the uniform integrability of the sequences $\{gX_{i,n+1}^2\}_{t=1}^{\infty}$ and $\{gY_{j,n+1}\}_{t=1}^{\infty}$ $(i = 1, \ldots, q, j = 1, \ldots, p)$ $\forall \ g \in G$, w.r.t. $\bar{\alpha}_t$, we have

$$D_{ii'} \to \frac{1}{k}\sum_{\ell=1}^{k} g_\ell(X'X)$$

and

$$A_{ij} \to \frac{1}{k}\sum_{\ell=1}^{k} g_\ell(X'Y) \ \text{ as } \ t \to \infty \ .$$

Thus, the limiting Bayes estimator of B (as $\alpha_t(R^d) \to 0$) is given by

$$B^{**} = \left(\sum_{\ell=1}^{k} g_\ell(X'X) \right)^{-1} \left(\sum_{\ell=1}^{k} g_\ell(X'Y) \right) . \quad (3.9)$$

If G has a single element, then from (3.6) and (3.9) we have

$$B^{*\alpha} = \left\{ p_n E^{\bar{\alpha}} \left(\boldsymbol{x}_{n+1} \boldsymbol{x}'_{n+1} \right) + (1 - p_n) \frac{1}{n} (X'X) \right\}^{-1} \times \left(p_n E^{\bar{\alpha}} \left(\boldsymbol{x}_{n+1} \boldsymbol{y}'_{n+1} \right) + (1 - p_n) \frac{1}{n} (X'Y) \right) , \tag{3.10}$$

and

$$B^{**} = (X'X)^{-1}(X'Y) . \tag{3.11}$$

The estimators $B^{*\alpha}$ and B^{**} given by (3.10) and (3.11), respectively are the Bayes and the limiting Bayes estimators of B under Dirichlet process prior $D(\alpha)$. Note that B^{**} coincides with the ordinary least squares estimator $\hat{B}$ given by (1.2).

When $G_{\boldsymbol{\mu}} = \{e, g_{\boldsymbol{\mu}}\}$, where $e(\boldsymbol{z}) = \boldsymbol{z}$, $g_{\boldsymbol{\mu}}(\boldsymbol{z}) = 2\boldsymbol{\mu} - \boldsymbol{z}$, then using the predictive distribution $\hat{F}_{\boldsymbol{\mu}}$ of $\boldsymbol{z}_{n+1}$ given in (2.4) we have

$$D_{ii'} = E^{\boldsymbol{\mu}|Z} \Bigg(p_n E^{\bar{\alpha}_\mu} \left(x_{i,n+1} \cdot x_{i',n+1} \right) + (1 - p_{n,\boldsymbol{\mu}}) \frac{1}{2n} \Big[\sum_{m=1}^{n} x_{im} x_{i'm} + \sum_{m=1}^{n} (2\mu_i - x_{im})(2\mu_{i'} - x_{i'm}) \Big] \Bigg) \qquad i, i' = 1, \ldots, q \tag{3.12}$$

and

$$A_{ij} = E^{\boldsymbol{\mu}|Z} \Bigg(p_{n,\boldsymbol{\mu}} E^{\bar{\alpha}_\mu} \left(x_{i,n+1} \cdot y_{j,n+1} \right) + \left(1 - p_{n,\boldsymbol{\mu}}\right) \frac{1}{n} \Big[\sum_{m=1}^{n} x_{im} y_{jm} + \sum_{m=1}^{n} (2\mu_i - x_{im})(2\mu_j - y_{jm}) \Big] \Bigg) \quad i = 1, \ldots, q , \quad j = 1, \ldots, p . \tag{3.13}$$

If $M_{\boldsymbol{\mu}}$ is independent of $\boldsymbol{\mu}$, we can replace $p_{n,\boldsymbol{\mu}}$ in (3.12) and (3.13) by p_n. Expressions (3.12) and (3.13) can be explicitly computed under the following cases.

(I) Diffuse prior for $\boldsymbol{\mu}$: Let $\bar{\alpha}$ have density given by $N_d(\mathbf{0}, I_d)$, the d-dimensional multivariate normal distribution with mean $\mathbf{0}$

and covariance matrix I_d. Then the density of $\bar{\alpha}\boldsymbol{\mu}$ is the density of $N_d(\boldsymbol{\mu}, I_d)$. Let ϕ denote the density of a standard normal distribution. Let $\boldsymbol{\mu}$ have a uniform prior on R^d with density $h(\boldsymbol{\mu}) = 1$, $\boldsymbol{\mu} \in R^d$. Then

$$E^{\bar{\alpha}\mu}(x_{i,n+1} \cdot x_{i',n+1}) = \mu_i \mu_{i'} \ ,$$
$$E^{\bar{\alpha}\mu}(x_{i,n+1} \cdot y_{j,n+1}) = \mu_i \mu_j \ , \quad 1 \le i, i' \le q \ , \ 1 \le j \le q \ . \tag{3.14}$$

If all the observations $\{ x_{ir}, y_{ir} \ ; \ 1 \le i \le q \ , \ 1 \le j \le p \ , \ 1 \le r \le n \}$ are assumed to be distinct, then from Dalal (1975, equation (32)) we have

$$\hat{\mu}_i = E^h(\mu_i \mid Z)$$
$$= \frac{a \prod_{r=1}^{n} \phi(x_{ir} - \bar{x}_i)\bar{x}_i + \sum_{1 \le r \ne s \ne t \le n} \phi(x_{it} - \frac{(x_{ir}+x_{is})}{2})\phi(\frac{(x_{ir}-x_{is})}{2})\frac{(x_{ir}+x_{is})}{2}}{a \prod_{r=1}^{n} \phi(x_{ir} - \bar{x}_i) + \sum_{1 \le r \ne s \ne t \le n} \phi\left(x_{it} - \frac{(x_{ir}+x_{is})}{2}\right)\phi\left(\frac{x_{ir}-x_{is}}{2}\right)} \tag{3.15}$$

for $1 \le i \le q$, and $a = 4M\sqrt{2\pi/n}$. The expression for $\hat{\mu}_j = E^h(\mu_j \mid Z)$, $1 \le j \le p$, is obtained from (3.15) by replacing x_{ir}, x_{is}, x_{it}, $\bar{x}_i$ by y_{jr}, y_{js}, y_{jt}, $\bar{y}_j$ respectively. From (3.12)–(3.15) we have

$$\begin{aligned} D_{ii'} &= p_n \hat{\mu}_i \hat{\mu}_{i'} + (1 - p_n)\frac{1}{nk}\left(\sum_{m=1}^{n} x_{im} x_{i'm} \right.\\ &\qquad \left. + \sum_{m=1}^{n} (2\hat{\mu}_i - x_{im})(2\hat{\mu}_{i'} - x_{i'm})\right) \\ &= (2 - p_n)\hat{\mu}_i \hat{\mu}_{i'} - (1 - p_n)[\hat{\mu}_i \bar{x}_{i'} + \hat{\mu}_{i'} \bar{x}_i] \\ &\qquad + \frac{(1 - p_n)}{2n} \sum_{m=1}^{n} x_{im} x_{i'm} \ , \end{aligned}$$
$$1 \le i \ne i' \le q \ , \tag{3.16}$$

and

$$A_{ij} = (2 - p_n)\hat{\mu}_i \hat{\mu}_j - (1 - p_n)\left[\hat{\mu}_i \bar{y}_j + \hat{\mu}_j \bar{x}_i\right] + \frac{(1 - p_n)}{2n} \sum_{m=1}^{n} x_{im} x_{jm} \ , \tag{3.17}$$
$$1 \le i \le q, \qquad 1 \le j \le p \ .$$

(II) Informative prior for $\boldsymbol{\mu}$. Suppose $\bar{\alpha}\,\boldsymbol{\mu} = N_d(\boldsymbol{\mu}, \sigma^2 I_d)$, $\sigma^2 > 0$ (known), and $\boldsymbol{\mu}$ has prior $N_d(\boldsymbol{\nu}, \boldsymbol{\tau}^2 I_d)$. In this case

$$\hat{\mu}_i = \frac{1}{D_i}\Bigg[a(x_i')\sum_{r=1}^{n}\phi\left(\frac{x_{ir} - \overline{\overline{x}}_i/b*}{(b*)^{-\frac{1}{2}}}\right)(\overline{\overline{x}}_i)^{\frac{1}{2}} \tag{3.18}$$

$$\times \sum_{1\le r\ne s\ne t\le n}\frac{1}{\sigma^2}\phi\left(\frac{1}{\sigma}\left(x_{it} - \frac{x_{ir}+x_{is}}{2}\right)\right)\phi\left(\frac{x_{ir}-x_{is}}{\sigma}\right)\left(\frac{x_{ir}+x_{is}}{2}\right)\Bigg]$$

where

$$b* = \left(\frac{n}{\sigma^2} + \frac{n}{\tau^2}\right), \qquad \overline{\overline{x}}_i = \left(\frac{n}{\sigma^2}\overline{x}_i + \frac{\nu_i}{\tau^2}\right)$$

and, for $1 \le i \le q$,

$$a(x_i') = 4M\int_{-\infty}^{\infty}\prod_{r=1}^{n}\frac{1}{\sigma}\phi\left(\frac{x_{ir} - \mu_i}{\sigma}\right)\frac{1}{\tau}\phi\left(\frac{\mu_i - \nu_i}{\tau}\right)d\mu_i \ ,$$

$$x_i = (x_{i1}, \dots x_{in})' \ ,$$

$$D_i = \Bigg[a(x_i')\sum_{r=1}^{n}\phi\Big(\frac{x_{ir} - \overline{\overline{x}}_i/b}{b^{-\frac{1}{2}}}\Big) \tag{3.19}$$

$$+ \sum_{1\le r\ne s\ne t\le n}\frac{1}{\sigma^2}\,\phi\Big(\frac{1}{\sigma}\big(x_{it} - \frac{x_{ir}+x_{is}}{2}\big)\Big)\phi\Big(\frac{x_{ir}-x_{is}}{\sigma}\Big)\Bigg] .$$

A similar expression for $\hat{\mu}_j$, $1 \le j \le p$, can be obtained from (3.18) and (3.19) by replacing x_{ir} , x_{is} , $\overline{x}_i$ and $\overline{\overline{x}}_i$ by y_{jr} , y_{js} , $\overline{y}_j$ and $\overline{\overline{y}}_j$, respectively. Now, D_{ii} , and A_{ij} can be obtained from (3.16)–(3.19).

The cases where the observations are not all distinct can be dealt with similarly. But the computations become very tedious and hence are omitted.

3.2 Bayes estimation with a mixture of Dirichlet invariant priors

Once again let $G = \{g_1, \dots, g_k\}$ be a finite group of measurable transformations from R into itself. For $\boldsymbol{z} \in R^d$, define $g\boldsymbol{z} = (g\boldsymbol{z}_1, \dots, g\boldsymbol{z}_d)\ \forall g \in G$. Then from (2.7), we have

$$D_{ii'} = p_n E^{u|Z} E^{\overline{\alpha}(u,\cdot)}\{x_{i,n+1}\cdot x_{i',n+1}\}$$

$$+ \frac{(1-p_n)}{nk}\sum_{m=1}^{n}\sum_{\ell=1}^{k} g_\ell(x_{i,n+1}\cdot x_{i',n+1}) \ , \quad 1 \le i\ , i' \le q \ , \tag{3.20}$$

and

$$A_{ij} = p_n E^{u|Z} E^{\overline{\alpha}(u,\cdot)|u} \{x_{i,n+1} \cdot y_{j,n+1}\} + \frac{(1-p_n)}{nk} \sum_{m=1}^{n} \sum_{\ell=1}^{k} g_\ell (x_{i,n+1} \cdot y_{j,n+1}) , \quad (3.21)$$

$$1 \le i \le q , \quad 1 \le j \le p .$$

The Bayes estimator of the regression coefficient matrix B can be obtained from (3.20) and (3.21).

We now provide an example in which by choice of elements of G, the Bayes estimates of B with a G-invariant mixture of Dirichlet processes prior reduce to the Bayes estimate with a mixture of Dirichlet processes prior. Let $G = \{e, g\}$ where $e(z) = z$ and $g(z) = -z$ for $z \in R$. Assume H, the mixing distribution, has density $h(U) = W_d^{-1}(U \mid C, \nu_0)$, $\nu_0 > 2d$, where W_d^{-1} denotes the density of an inverted Wishart distribution with positive definite matrix C and $f = \nu_0 - d - 1$ degrees of freedom. Given U, let $\alpha(U, \cdot) = N_d(\cdot \mid \mathbf{0}, U)$ be the transition measure. Note that $\alpha(U, \cdot)$ is G-invariant. Thus, the assumptions of Theorem 2.5 hold. As shown in Poli (1985) the posterior density of the random matrix U is $h(U \mid Z) = W_d^{-1}(U \mid G, \nu*)$, where $G = ((YX)'(YX) + C)$, and $\nu* = n + \nu_0$. Thus, from (3.20) and (3.21) we have

$$D_{ii'} = p_n \int_{R^d} x_i x_i' \, dS_d(\boldsymbol{y}, \boldsymbol{x} \mid \mathbf{0}, H_{\nu*}, \nu*) + \frac{(1-p_n)}{n} \sum_{m=1}^{n} x_{im} \cdot x_{i'm}, \quad 1 \le i, \; i' \le q , \quad (3.22)$$

and

$$A_{ij} = p_n \int_{R^d} x_i y_j \, dS_d(\boldsymbol{y}, \boldsymbol{x} \mid \mathbf{0}, H_{\nu*}, \nu*) + \frac{(1-p_n)}{n} \sum_{m=1}^{n} x_{im} y_{jm}, \quad 1 \le i \le q, \; 1 \le j \le p , \quad (3.23)$$

where S_d is the distribution function of a d-variate student distribution with mean vector $\mathbf{0}$ and covariance matrix $H_{\nu*} = (\nu^*) G^{-1}$.

The Bayes estimator of B can now be derived using (3.22) and (3.23), and is identical to Poli (1985), equation (6).

REFERENCES

Antoniak, C. E. (1974), "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems". *Annals of Statistics* **2,** 1152–1174.

Dalal, S. R. (1975), "Contributions to Bayes nonparametric decision theory". Ph.D. Dissertation, University of Rochester, Department of Statistics.

Dalal, S. R. (1979), "Dirichlet invariant processes with applications to nonparametric estimation of symmetric distribution functions". *Stochastic Processes and their Applications* **9,** 99–107.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems". *Annals of Statistics* **1,** 209–230.

Goldstein, M. (1976), "Bayesian analysis of regression problems". *Biometrika* **63,** 51–58.

Parthasarathy, K. R. (1967), *Probability Measures on Metric Spaces.* Academic Press.

Poli, I. (1985), "A Bayesian nonparametric estimator for multivariate regression". *Journal of Econometrics* **28,** 171–182.

Prakasa Rao, B. L. S. (1983), *Nonparametric Functional Estimation.* Academic Press.

Press, S. J. (1982), *Applied Multivariate Analysis: Using Bayesian and Frequestist Methods of Inference.* Robert E. Krieger Publishing Co.

Sethuraman, J., and Ram C. Tiwari (1982), "Convergence of Dirichlet measures and the interpretations of their parameter". *Statistical Decision Theory and Related Topics*, III, Vol. 2. (S. S. Gupta and James O. Berger, Eds.). Academic Press.

Tiwari, Ram C. (1981), "A mathematical study of the Dirichlet process". Ph.D. Dissertation, Florida State University, Department of Statistics.

Wilks, S. S. (1962), *Mathematical Statistics*, 2nd ed. New York: John Wiley.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley.